APPLICATION OF DISCRIMINATE ANALYSIS TO PREDICTION OF COMPANY FUTURE ECONOMIC DEVELOPMENT

VYUŽITÍ DISKRIMINAČNÍ ANALÝZY PRO PREDIKCI BUDOUCÍHO VÝVOJE FIRMY

Radmila SOUSEDÍKOVÁ [1], Jaroslav DVOŘÁČEK [2], Igor SAVIČ [3]

[1] *RNDr. Ph.D., Institute of Economics and Control Systems, Faculty of Mining and Geology, VŠB–Technical University of Ostrava*
17. listopadu 15, Ostrava-Poruba, tel. (+420) 59 732 3850
e-mail: radmila.sousedikova@vsb.cz

[2] *prof. Ing. CSc., Institute of Economics and Control Systems, Faculty of Mining and Geology, VŠB–Technical University of Ostrava*
17. listopadu 15, Ostrava-Poruba, tel. (+420) 59 732 3381
e-mail: jaroslav.dvoracek@vsb.cz

[3] *Mgr., Cybex Industrial LTD, C/O Columbus Trading-Partners GMBH*
Riedinger str. 18, 95448 Bayreuth, German, tel. (+49) 921 78 511 273
e-mail: igor.savic@cybex-online.com]

Abstract

The paper takes into account applications of discriminate analysis as regards prediction of future economic development of companies. An assumption of multivariate normality of discriminators has been tested and outliers identified. An outlier reduction of original data files brings data distribution closer to multivariate normality, and substantially improves discriminate function classification abilities**.**

Abstrakt

Článek se zabývá využitím diskriminační analýzy pro predikci budoucího vývoje firmy. Je testován předpoklad vícerozměrné normality diskriminátorů a jsou identifikovány vybočující hodnoty. Redukce původních datových souborů o vybočující data přispívá k přiblížení rozložení dat vícerozměrné normalitě a vede k podstatnému zlepšení klasifikační schopnosti diskriminační funkce.

Key words: Default prediction, discriminate analysis, multivariate normality, outliers

1 INTRODUCTION

The initial comprehensive studies of modelling future development of companies date back to the thirties of the twentieth century. These works take into account analyses of financial ratios as regards successful and failed companies. Ramser and Foster (1931) compared ratios for successful and failed businesses and could prove that corporate finance of successful firms demonstrate better ratio figures than it would be the case with those in distress or bankrupt. What more, Fitzpatrick (1932) tried to identify those corporate finance ratios that might serve the purpose of predicting future economic developments. Smith and Winakor (1935) analyzed book keeping records of bankrupt companies, identifying indices of imminent default.

A study of cardinal importance is that of Beaver (1966) who analyzed the utilisation of financial ratios for predicting future economic development of companies. Beaver designed his study to be a benchmark for future investigations into alternative predictors of failure. He succeeded in providing an empirical verification of the predictive ability of financial statement accounting data, namely the financial ratios, for the prediction of corporate business failure.

All the studies that have been mentioned above demonstrate the ability of financial ratios to predict bankruptcy, and they try to identify a single ratio or a group of such ratios whose values would indicate imminence of default. Nevertheless, conclusions of these studies are not unambiguous, because each study exemplifies different financial ratios as indicators of default imminence. The period from the thirties to the end of the sixties of the 20th century can be qualified as the period of one-dimensional analyses of financial ratios, when the individual ratios are analyzed in their isolation and their interaction is disregarded.

Altman (1968) built on the existing results and could overcome limitations of one-dimensional approaches by combining several financial ratios for his developments of prediction models. He solved problems of the capability of specific ratios to predict default. He could provide for objective quantification of each specific ratio by defining weights that should be attached to these selected ratios and how these weights should be objectively established. Altman is innovative in his utilization of the multiple discriminate analyses by developing a multi-dimensional model of a corporate business default. In his later studies, Altman adapted his model several times and along with Halderman and Naryanan (1977) developed a second generation of prediction models for corporate business failure. Altman is aware of deficiencies of models that utilize linear discriminate analysis (Altman et. al., 1981), especially as regards non-conformity to the assumption multivariate normal distribution of discriminators in each class, and equality of class covariance matrices.

The impossible fulfilment of multidimensional normality resulted in the utilization of other distribution functions that would describe the input data distribution. Ohlson (1980) was the first one who used a logistic regression analysis for creating prediction models of corporate business defaults.

Since the beginning of the nineties, studies have been performed that utilize neuron networks for modelling of corporate failures (Odom and Sharda, 1990; Tam, 1991).

2 DISCRIMINATE ANALYSIS

The discriminate analysis enables the evaluation of differences between two or more subject groups that are characterised by a certain number of features. Such evaluation provides a base of classification that builds on it. If the discriminate analysis is applied for predicting future economic footing of companies, the subjects are particular firms that are structured into two groups of prosperous firms and those in jeopardy of default. Each firm is characterised by a definite number of quantitative variables called discriminators.

The choice of discriminators plays a major role as regards prediction abilities of final modelling. The discriminator calculation data can be drawn from common accounting documentation – Balance Sheet, Profit and Loss Statement – and it provides for calculations of ratios or indexes.

There are two ways for choosing the discriminators. The first way takes advantage of a mathematical approach to the problem, and takes a lot of various ratios and indexes into account, and so we are not sure which of them are effective for classifying a particular firm as prosperous or threatened by default. We can distinguish particular discriminator efficiencies by assessing the changing values of the Mahalanobis distance, $D_M{}^2$, between mean values of both classes, that have been effected by adding or removing a definite discriminator. The individual steps of discriminator addition or removal can be directed by various decision-making criteria (for example the Wilks' Criterion, λ).

The second way has been based on experience, knowledge, and intuition of researchers, when the choice of discriminators is supported by a theoretical model for solving the given task. This paper's investigation has opted for this second way and the discriminator selection has been based on the assumption that bankruptcy is caused by disrupted circulation of capital. That is why, 8 financial ratios were chosen, namely,

1. Quick liabilities/Total assets
2. Current assets index
3. Current liquid assets/Current assets
4. Sales index
5. Total assets index
6. Receivables/Current assets
7. Index, Loan capital/Total assets
8. Equity capital/Total assets

2.1. Model

The discriminate analysis model consists in a linear combination of variables, so called discriminators that best distinguish between prosperous and default companies.

This is the linear discriminate function formula:

$$D_i = d_1 X_{i1} + d_2 X_{i2} + \cdots + d_m X_{im} , \tag{1}$$

where

n – Number of firms in the class,

m – Number of discriminators,

D_i – Discriminate score for firm, i,

X_{ij} – Discriminator value for firm, i, ($j = 1,\ldots,m$),

d_j – Linear discriminate coefficient for discriminator, j, (for $j = 1,\ldots,m$).

This formula combines several firm's characteristics (discriminators) into a single multivariate score, D_i , whose value is between $-\infty$ and ∞ , and indicates financial health of the firm. The discriminate score low value, D_i, marks bad financial health of the firm.

A correct application of the linear discriminate analysis asks for observation of the following requirements:

- Discriminators evidence multivariate normality of distribution,
- Classes of prosperous firms and those in jeopardy of default have the same covariance matrices.

2.2. Testing normality of discriminators

In the following, we concentrate on multivariate normality of discriminators. The multivariate normality of independent variables (discriminators) should be verified by an appropriate statistical test. The test of multivariate normality is difficult. Although the random vector particular constituents of all discriminators evince one-dimensional normality, the associated density of probability does not necessarily have multivariate normal distribution. As such, the one-dimensional normality of particular discriminators is a necessary but insufficient prerequisite of vector multivariate normality of all discriminators. This fact can serve as a tool of practical verification of multivariate normality of discriminators. The one-dimensional normality of particular discriminators is verified first. If, at least, one discriminator does not evidence one-dimensional normality, it is obvious that no multivariate normality exists, and the testing is terminated. If all features evidence normality, it is necessary to continue in testing of multivariate normality.

For example, the normality of particular discriminators can be verified by tests of skewness and kurtosis.

The testing criterion is defined by the following formula [6]:

$$C_1 = \frac{\hat{g}_1^{\,2}}{D(\hat{g}_1)} + \frac{[\hat{g}_2 - E(\hat{g}_2)]^2}{D(\hat{g}_2)}, \tag{2}$$

where

n - Sample size,

x_i - Value of the discriminators tested for firm, i,

$\bar{x}$ - Mean value of the discriminator tested,

$$\hat{g}_1 = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{3/2}} \quad \text{- Sample skewness,} \tag{3}$$

$$D(\hat{g}_1) = \frac{n-2}{(n+1)(n+3)} \quad \text{- Variance of sample skewness,} \tag{4}$$

$$\hat{g}_2 = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{2}} \quad \text{- Sample kurtosis} \tag{5}$$

$$E(\hat{g}_2) = 3 - \frac{6}{n+1} \quad \text{- Expected value of sample kurtosis,} \tag{6}$$

$$D(\hat{g}_2) = \frac{24\,n\,(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} \quad \text{- Variance of sample kurtosis.} \tag{7}$$

If normality is evidenced, the value, C_1 , has an asymptotic distribution, χ^2 , with 2 degrees of freedom. If it is verified that $C_1 > \chi^2_{1-\alpha}(2)$, it is necessary to reject hypotheses that the sample is normally distributed [6].

The normality of particular discriminators can be also assessed graphically by means of rankit plots, which enable comparison of each discriminator distribution with normal distribution. The rankit plots are executed by plotting the quantiles of the standard normal distribution, u_{p_i}, on the plot's horizontal axis, and the order statistics, $x_{(i)}$ (discriminator values structured hierarchically in ascending order), on its vertical one. If a discriminator's distribution equals normal distribution, we can observe a linear dependence, $x_{(i)}$, on u_{p_i}.

For the class of prosperous firms, as well as for those in jeopardy of default, the tests of one-dimensional normality for all eight discriminators were performed. The value of the testing criterion, C_1, for particular discriminators (see Tab. 1), and their comparisons with the quantile, $\chi^2_{0,95}(2)=5{,}99$, make it obvious that none of the discriminator distributions is normal.

Tab. 1 Test criterion values for the normality verification of particular discriminators

Discriminator	Test criterion value, C_1	
	Prosperous firms	Firms in jeopardy of default
Quick liabilities/Total assets	206.39	4941.70
Current assets index	10146.11	138.12
Current liquid assets/Current assets	61.94	869.67
Sales index	10661.24	4890.87
Total assets index	2563.03	6.47
Receivables/Current assets	2325.0	24.37
Index, Loan capital/Total assets	104.05	12035.42
Equity capital/Total assets	11.64	4019.53

Also rankit plots of all discriminators for both classes of firms evidence the fact that no discriminator distribution is normal. The following Figs. 1, 2, provide for a comparison of rankit plots of sample discriminators for the class of profitable businesses. The distribution of the discriminator, Equity capital/Total assets (see Fig. 1) approximates the normal distribution most closely. In contrast to this, the distribution of the discriminator, Sales index (see Fig. 2) evidences the worst result as regards the distribution normality.

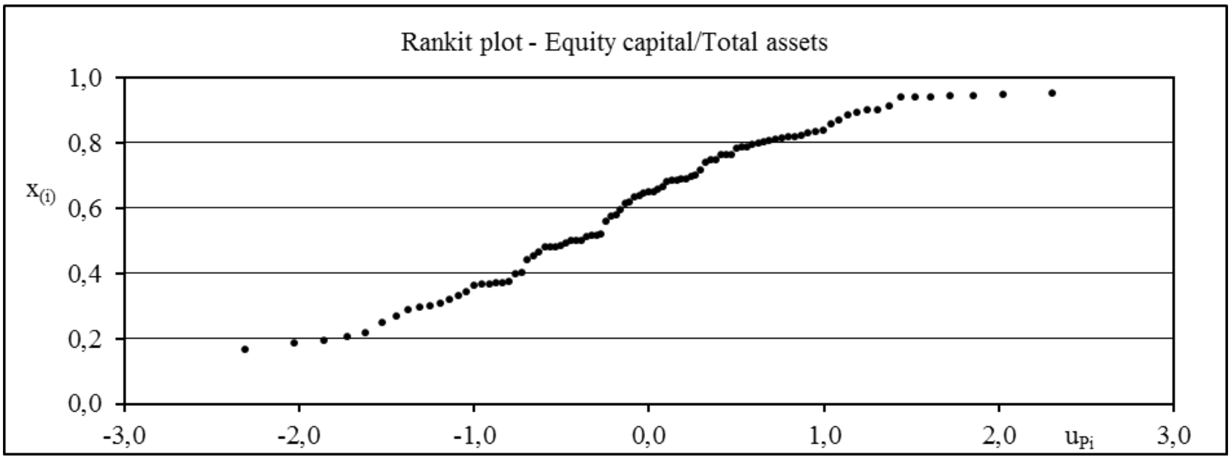

***Fig. 1** Rankit plot for the Equity capital/Total assets (prosperous firms).*

Volume LVIII (2012), No.4
p. 1-9, ISSN 1802-5420

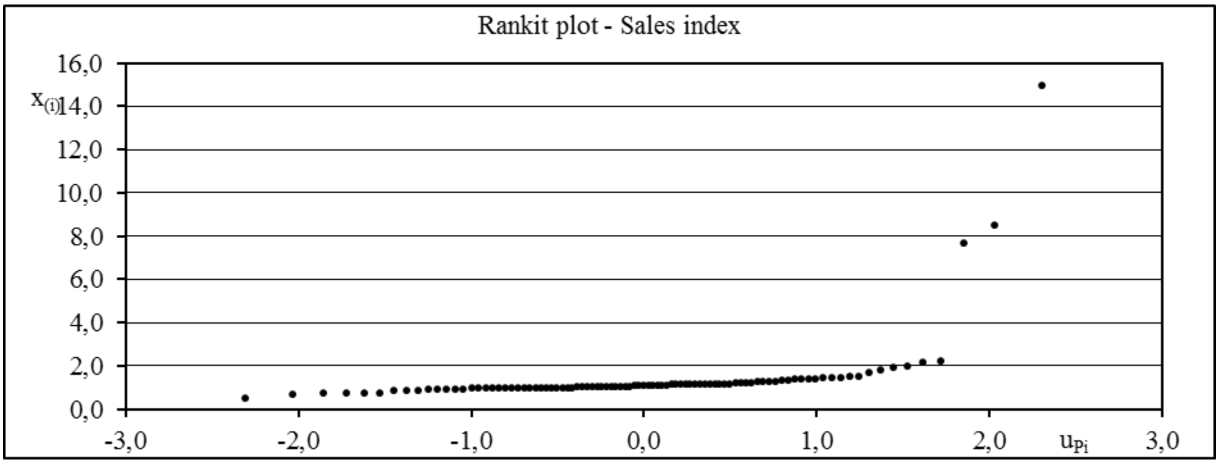

Fig. 2 Rankit plot for the Sales index (prosperous firms).

Analogically, for the class of firms in jeopardy of default, Fig. 3 provides for the rankit plot of the discriminators, Total assets index, which shows almost a linear dependence close to normality. In contrast to this, Fig. 4 testifies that the rankit plot for the discriminator, Index, Loan capital/Total assets, evidences a pronounced deviation from the normal distribution

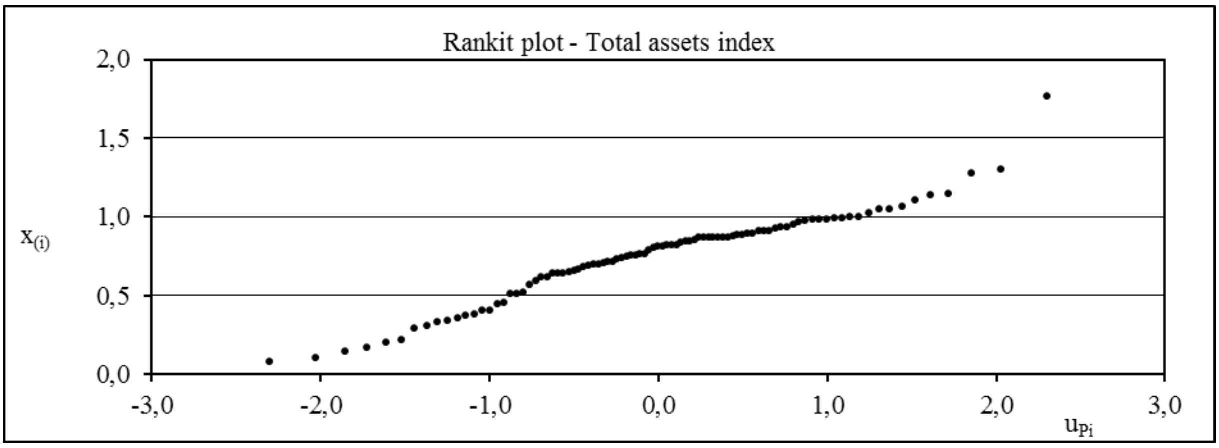

Fig. 3 Rankit plot for the Total assets index (default firms).

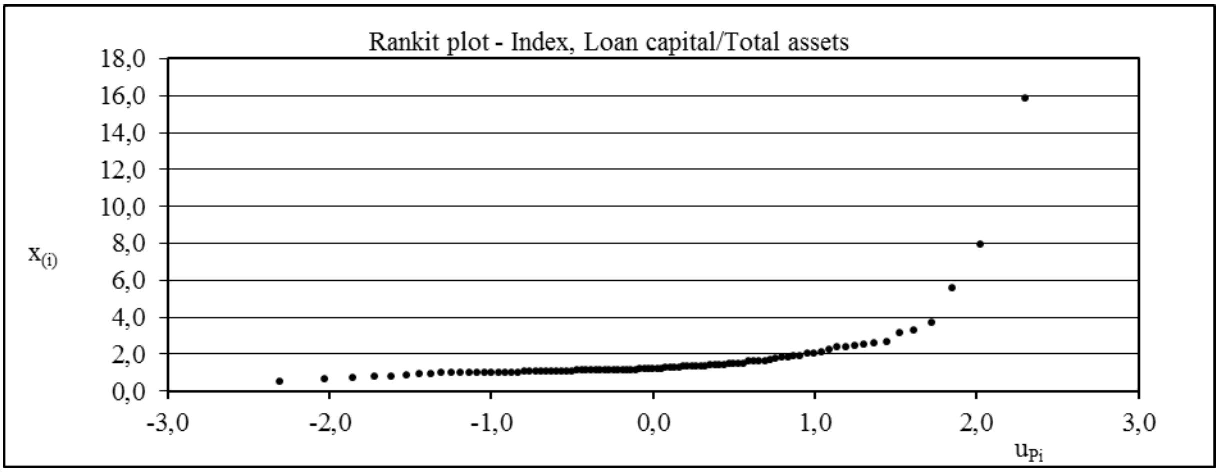

Fig. 4 Rankit plot for the Index, Loan capital/Total assets (default firms).

The above given data are not surprising. For the economic data, a normality deviation is rather the rule than the exception. The majority of research works (inclusive Altman) do not test multivariate normality and assume that models are sufficiently robust for providing rational approximations even without meeting the prerequisite of input data normal distribution.

Having this in mind, let us try to improve the data distribution by exclusion of outliers, and let us compare the results of the discriminate analysis before and after this modification of the data.

2.3. Identification of outliers

Two approaches will be taken for the outlier identification:

1. One-dimensional approach: Particular discriminator outliers are identified by so called inner fences.
2. Multivariate approach: Outliers are identified by their Mahalanobis distance from the mean value data.

Outlier identification by inner fences:

The outliers are all discriminator values that lie outside the interval,

$$\left(B_D^*;B_H^*\right), \tag{8}$$

$$B_D^* = \tilde{x}_{25} - K\left(\tilde{x}_{75} - \tilde{x}_{25}\right), \tag{9}$$

$$B_D^* = \tilde{x}_{75} - K\left(\tilde{x}_{75} - \tilde{x}_{25}\right), \tag{10}$$

$$K = 2{,}25 - \frac{3{,}6}{n}, \tag{11}$$

where

n - Sample size,

$\tilde{x}_{25}$ - Lower quartile,

$\tilde{x}_{75}$ - Upper quartile [6].

The following Tab. 2 gives numbers of discriminator outliers for both classes of firms.

Tab. 2 Number of discriminator outliers

Discriminator	Number of outliers	
	Prosperous firms	Default firms
Quick liabilities/Total assets	2	5
Current assets index	8	1
Current liquid assets/Current assets	0	9
Sales index	7	3
Total assets index	1	1
Receivables/Current assets	1	0
Index, Loan capital/Total assets	4	6
Equity capital/Total assets	0	7

Firms that have one or more discriminator outliers were excluded from the investigation. The original samples of 93 prosperous and 93 default companies were reduced to 76 prosperous and 74 default companies.

The outlier identification by the Mahalanobis distance:

The outliers are all multivariate data of the formula,

$$d_i^2 = \left(\mathbf{x_i} - \bar{\mathbf{x}}\right)^T \mathbf{S}^{-1}\left(\mathbf{x_j} - \bar{\mathbf{x}}\right) > \chi^2_{1-\frac{\alpha}{n}}(m), \tag{12}$$

$\mathbf{x_j}$ - Discriminator vector of particular firms $(j = 1,\ldots,n)$,

$\bar{\mathbf{x}}$ - Sample mean vector of particular discriminators,

$\mathbf{S}^{-1}$ - Inverse co-variance matrix,

$\chi^2_{1-\frac{\alpha}{n}}(m)$ - Quantile of distribution, χ^2, with degrees of freedom, m

where

m - Number of discriminators,

α - Significance level,

n - Number of firms of particular class [6].

The Mahalanobis distance identified five prosperous and eight default firms as having outlying discriminator values. As such, the original samples were reduced to 88 prosperous and 85 default companies.

The data distribution improvement can be demonstrated by a Q-Q plot. The plot illustrates dependence of order statistics, $C_{(i)}$ (C_i values structured hierarchically in ascending order) on values, C_i^*, where

$$C_i = \frac{n}{(n-1)^2}(\mathbf{x_i} - \bar{\mathbf{x}})^T \mathbf{S}^{-1}(\mathbf{x_j} - \bar{\mathbf{x}}), \tag{13}$$

$$C_i^* = \frac{i-a}{n-a-b+1}, \tag{14}$$

where

$$a = \frac{0{,}5m-1}{m},$$

$$b = \frac{0{,}5(n-m-1))-1}{n-m-1}.$$

In case of multivariate normality, the dependence of $C_{(i)}$ on C_i^* should be a linear one.

The Fig. 5 provides for the Q-Q plot of 93 prosperous firms

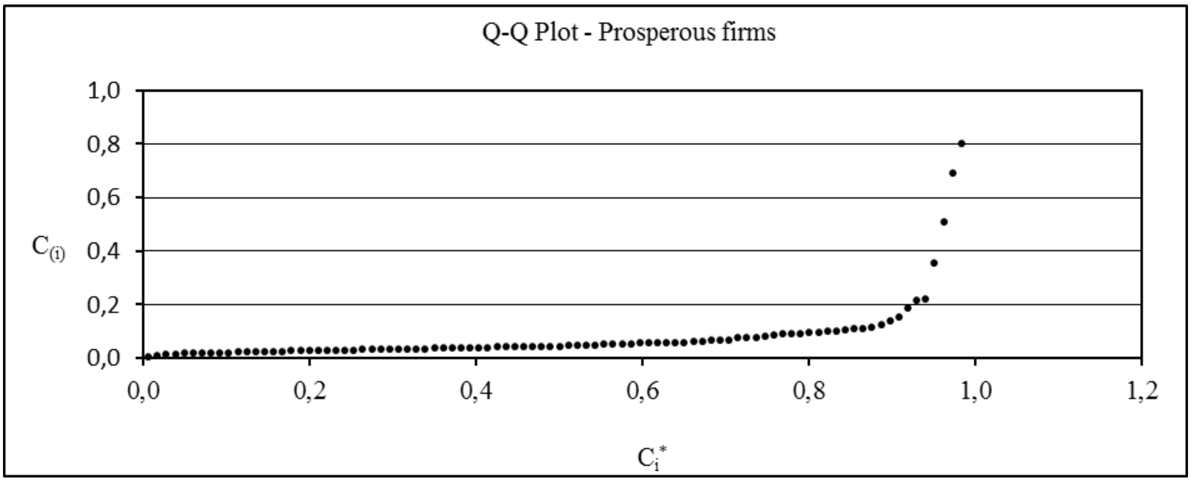

Fig. 5 Q-Q Plot of prosperous firms – Original sample of 93 firms

The following Fig. 6 gives the Q-Q Plot for the reduced sample of 76. The reduction was made by removing firms that had some discriminator values identified by inner fences as outliers.

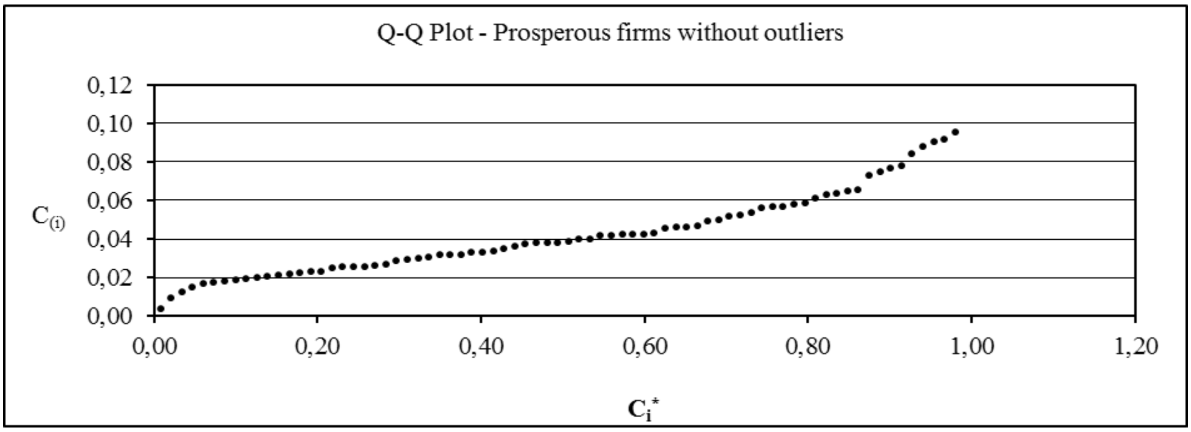

Fig. 6 Q-Q Plot of prosperous firms – Reduced sample of 76 firms

The comparison of these two plots makes it obvious that removal of outliers improved considerably the data distribution. The almost linear dependence of the plot in Fig. 6 provides for the possible conclusion that the reduced sample of prosperous firms closely approximates normal distribution.

3 DISCRIMINATE ANALYSIS OUTCOME COMPARISON

For the samples reduced by outlier elimination, the linear discriminate functions were calculated. These provided for the classification of firms as prosperous or in jeopardy of default. The results of this classification were compared with those, which had been the outcome of working with original non-reduced samples.

The comparison of results gives Tab. 3.

Tab. 3 Outcome comparison of working with original and reduced samples

	Classification success [%]		
	Prosperous firms	Bankrupt firms	Firms total
Original sample (93 prosperous firms + 93 in jeopardy of failure)	87.10	89.25	88.17
Sample reduced by inner fences (76 prosperous firms + 74 in jeopardy of failure)	100.00	90.54	95.33
Sample reduced by Mahalanobis distance (88 prosperous firms + 75 in jeopardy of failure)	95.45	94.12	94.80

It is obvious from the aforementioned results that approximation to multivariate normality improves the discriminate function classification abilities considerably. Concerning the approximation method specificity, i.e. the discriminate outlier elimination, the reduction by inner fences led to better results. Regarding the multivariate normality, the samples of which the outliers were eliminated by inner fences demonstrated better results which were corroborated by the Q-Q Plots.

4 CONCLUSION

The linear discriminate analysis represents a method of predicting future economic development of firms that is often applied. The choice of discriminators has a considerable impact on the prediction abilities of the final model. The selection of particular discriminators was based on the idea that bankruptcy is caused by disrupted circulation of capital. The correct application of discriminate analysis asks, apart from other requirements, for fulfilment of the prerequisite of the discriminator normal distribution. The majority of research that has been conducted in the field does not test the multivariate normality and assumes that models are sufficiently robust, providing for realistic results even without meeting the distribution normality requirements. The testing of the multivariate normality of discriminators was performed in the way, which first tested one-dimensional normality of discriminators by tests of skewness and kurtosis. The one-dimensional normality was also assessed by rankit plots. The testing of the aforementioned discriminators led to the conclusion that none of the discriminators tested had normal distribution, which was the reason why the requirement of multivariate normal distribution of all discriminators could not be met. Such results are not surprising because economic data often deviate from normal distribution. That was the reason why the data distribution was modified. The samples of prosperous and default firms were adapted in two ways. The one-dimensional way consisted in analysing particular discriminators and excluding the outlying discriminator values by inner fences. The multivariate way started with the elimination of outliers by the Mahalanobis distance. The improvement of the data distribution is testified by Q-Q Plots. The reduced samples were subject to a discriminate analysis which provided for the classification of firms as successful or failed. It is obvious from the results obtained that the approximation to multivariate normality improves considerably the classification abilities of the discriminate function.

REFERENCES

[1] ALTMAN, E. I. Financial ratios, discriminate analysis and prediction of corporate bankruptcy. *The Journal of Finance.* 1968, XXIII. Nr. 4, pp. 589-609.

[2] ALTMAN, E., HALDERMAN, R. & NARAYANAN, P. ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance.* 1977, I. Nr. 1, pp. 29-54.

[3] ALTMAN, E. I. et al. *Application of Classification Techniques in Business, Banking and Finance. (Contemporary Studies in Economic and Financial Analysis)*. Greenwich, Conn. : Jai Press, 1981. 418 pp.

[4] BEAVER, W. Financial Ratios as Predictors of Failure. *Journal of Accounting Research. 1966*, IV. Empirical Research in Accounting: Selected Studies 1966, pp. 71-111.

[5] FITZPATRICK, P. J. A Comparison of ratios of successful industrial enterprises with those of failed firms. *Certified Public Accountant*, 1932. October, November, December, pp. 598-605, pp. 656-662, pp. 727-731.

[6] Meloun, M. & Militký, J. *Statistická analýza experimentálních dat (Statistical Analysis of Experimental Data).* Praha : Academia, 2004. ISBN 80-200-1254-0.

[7] ODOM, M. D. a R. SHARDA. A Neural network model for bankruptcy prediction. *IJCNN International Joint Conference on Neural Networks.* 1990, I-III. pp. B163-B168.

[8] OHLSON, J. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*. 1980, XVIII. Nr. 1, pp. 109-131.

[9] RAMSER, J. R. & FOSTER, L. O. A Demonstration of Ratio Analysis. *Bulletin No. 40*. 1931, Bureau of Business Research, University of Illinois, Urbana.

[10] SMITH, R. F. and A. H. WINAKOR. *Changes in the financial structure of unsuccessful industrial corporations*. Urbana: University of Illinois, 1935.

[11] TAM, K. Y. Neural network models and prediction of bank bankruptcy. *Omega-International Journal of Management Science.* 1991, IXX. Nr. 5, pp. 429-445.

RESUMÉ

Článek se zabývá využitím diskriminační analýzy pro predikci budoucího vývoje firmy. Predikční schopnost výsledného modelu je významně ovlivněna volbou vstupních proměnných – diskriminátorů. Výběr použitých diskriminátorů byl založen na názoru, že příčinou bankrotu je porušení koloběhu kapitálu. Korektní aplikace diskriminační analýzy vyžaduje mimo jiné splnění předpokladu vícerozměrné normality rozdělení diskriminátorů. Většina výzkumných prací však vícerozměrnou normalitu netestuje a předpokládá, že modely jsou dostatečně robustní a dávají rozumné aproximace i bez splnění tohoto předpokladu. Negativní výsledky testování jednorozměrné normality jednotlivých diskriminátorů pomocí testu kombinace šikmosti a špičatosti vedly k závěru, že nemůže být splněna ani vícerozměrná normalita rozložení všech diskriminátorů. Proto bylo rozložení dat upraveno pomocí vyloučení vybočujících hodnot. Datové soubory zdravých firem i firem ohrožených bankrotem byly redukovány jednak pomocí vnitřních hradeb, jednak pomocí Mahalanobisovy vzdálenosti od středních hodnot. Na redukované soubory pak byla aplikována diskriminační analýza a provedena klasifikace firem na úspěšné a neúspěšné. Z výsledků klasifikace je zřejmé, že snaha o přiblížení rozložení dat vícerozměrné normalitě vede k podstatnému zlepšení klasifikační schopnosti diskriminační funkce.